Dec 22, 2020
SWELL Workshop

# The Processing of Korean Relative Clauses

Soyeon Kang, Gyuhwan Lee, and Sunwoo Jeong
Meanings & Melodies Lab

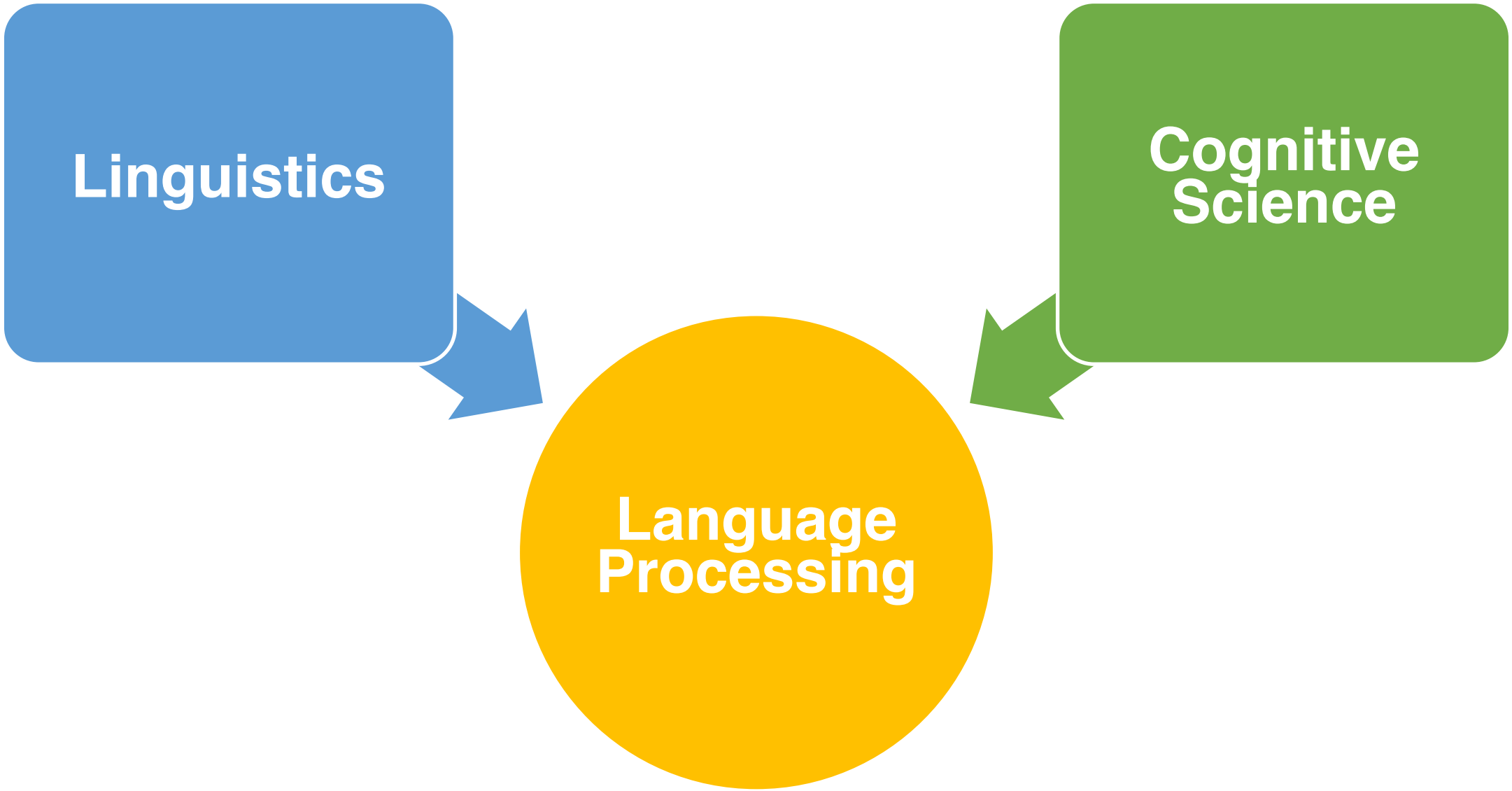
Massachusetts Institute of Technology

brain+cognitive sciences

생활과학대학 아동가족학과
Dept. Child Development & Family Studies
College of Human Ecology

# My Backgrounds

- Psycholinguistics
- Language Acquisition
- Syntax, and Pragmatics


Linguistics
Cognitive Science
Language Processing

# 1. English Relative Clauses

Asymmetry?

The reporter [who met the doctor]

The reporter [who the doctor met]

If so, why?

# 1.1. Relative Clauses (RCs)

- Contain a wh-pronoun that ‘relates’ to an antecedent in a higher clause.
- Long-distance dependency: an argument of a verb is displaced from its canonical position to a position in the sentence at some distance from the verb.

The reporter$_i$ [who ___$_i$ attacked the senator] met the editor.

filler gap

(See: e.g., Sprouse et al. (2016))

# 1.2. Asymmetry in RCs

- Noun Phrase Accessibility Hierarchy (Keenan & Comrie, 1977)
  Linguistic typology: Subject > Direct Object > Indirect Object >
  Oblique > Genitive > Object of Comparison

- SU: the boy$_i$ [who ___$_i$ loves the girl]
  DO: the girl$_i$ [who the boy loves ___$_i$]
  IO: the boy$_i$ [who the girl gives ___$_i$ a letter]
  OBL: the girl$_i$ [who the boy talks about ___$_i$]
  GEN: the boy$_i$ [whose ___$_i$ name is Tom]
  OCOMP: the girl$_i$ [who I am taller than ___$_i$]

# 1.3. 3 Hypotheses on the Asymmetry

| Hypotheses | Predictions in English |
|---|---|
| Working Memory Hypothesis | SR > OR |
| Syntactic Distance Hypothesis | SR > OR |
| Syntactic Probability Hypothesis | SR > OR |

# A > B: A is easier than B.

# 1.3.1. Working Memory Hypothesis (Gibson, 2000)

The processing difficulty of relative clause can be explained by the linear distance between the filler and the gap, which is measured by the number of intervening words between them.

(1) Linear distance in English RCs

a. Subject relative clause

the man [**who** _ loves the woman]: 1 word

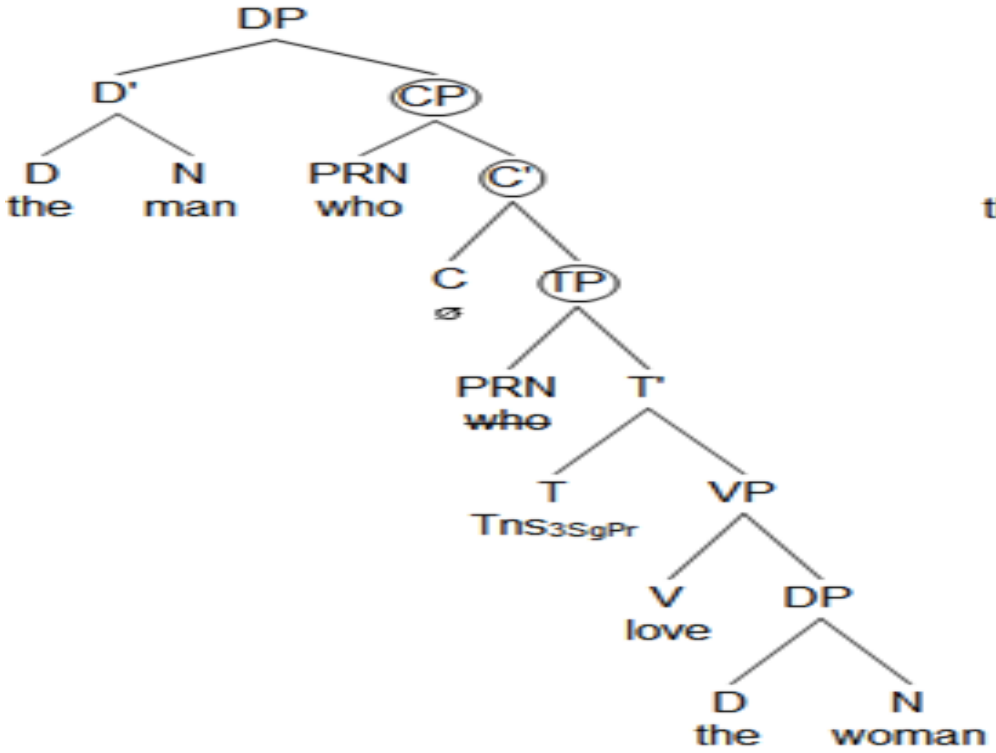
b. Object relative clause

the man [**who the woman loves** _ ]: 4 words

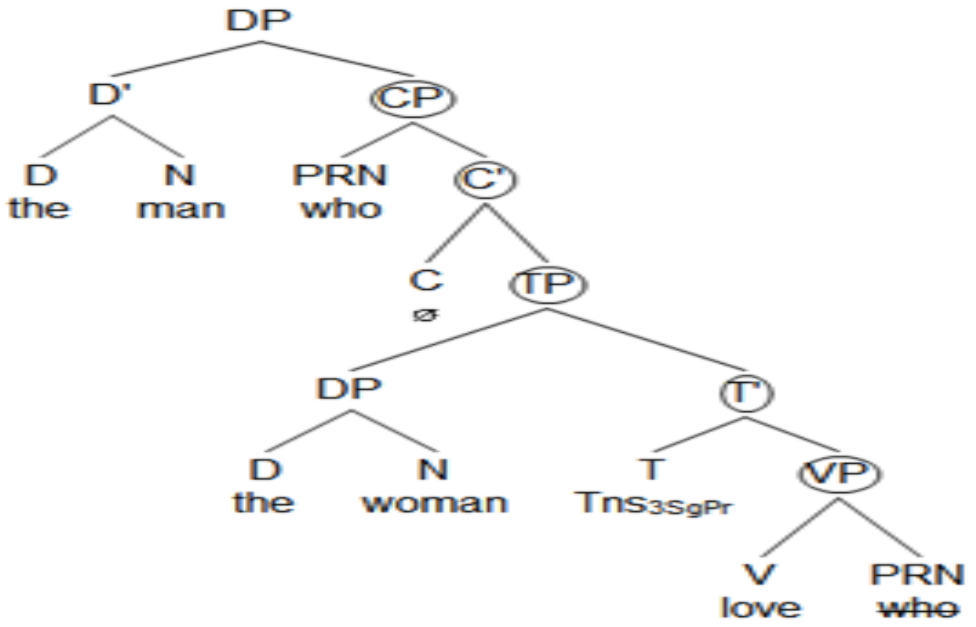
# 1.3.2. Syntactic Distance Hypothesis (O'Grady, 1997)

The processing difficulty of a relative clause is determined by the structural distance between the filler and the gap, which is calculated by the number of syntactic nodes intervening between them.

**SR:** the man [who _ loves the woman]




**OR:** the man [who the woman loves _]



# 1.3.3. Syntactic Probability Hypothesis

The processing difficulty of a relative clause is determined by probabilistic knowledge of structure.

(a) Entropy (Hale, 2006): the uncertainty about the rest of the sentence

(b) Surprisal (Levy, 2008): the expectation about upcoming structures

SR: the reporter [that attacked the senator] admitted the error.

OR: the reporter [<u>the senator</u> attacked] admitted the error.

less expected

# 2. Korean Relative Clauses

Same Asymmetry?

[기자를 만난] 의사

[기자가 만난] 의사

If so, why?

# 2.1. Asymmetry in Korean RCs

- Behavioral experiments (Kwon, 2008)

  : Eye-tracking, self-paced reading, and ERP

SR: [그 신문사의 편집장을 뇌물혐의로 협박한] 총장이 기자를 어제 만났다.
OR: [그 신문사의 편집장이 뇌물혐의로 협박한] 총장이 기자를 어제 만났다.

→ SR > OR

# 2.2. 2 Supported Hypotheses?

Behavioral experiments (Kwon, 2008)

| Hypotheses | Predictions in Korean |
|---|---|
| Working Memory Hypothesis | SR < OR |
| Syntactic Distance Hypothesis | SR > OR |
| Syntactic Probability Hypothesis | SR > OR |

# A > B: A is easier than B.

# 2.3. Confounding Variables?

My ongoing self-paced reading experiment

[Context]

야구경기 도중에 한 선수가 화가 나서 코치를 때렸다. 그러자 그 코치가 화가 나서 다른 선수를 때렸다.

윤아가 말했다: 나는 티파니가 그 선수들 중 한 명을 만났고, 제시카가 다른 한 명을 만났다고 들었어. 제시카가 만났던 선수는 누구일까?

[Target sentence]

유리가 말했다: 코치가 / 때렸던 /선수 / 인 것 / 같아.

코치를 / 때렸던 / 선수 / 인 것 / 같아.

# 2.3. Confounding Variables?

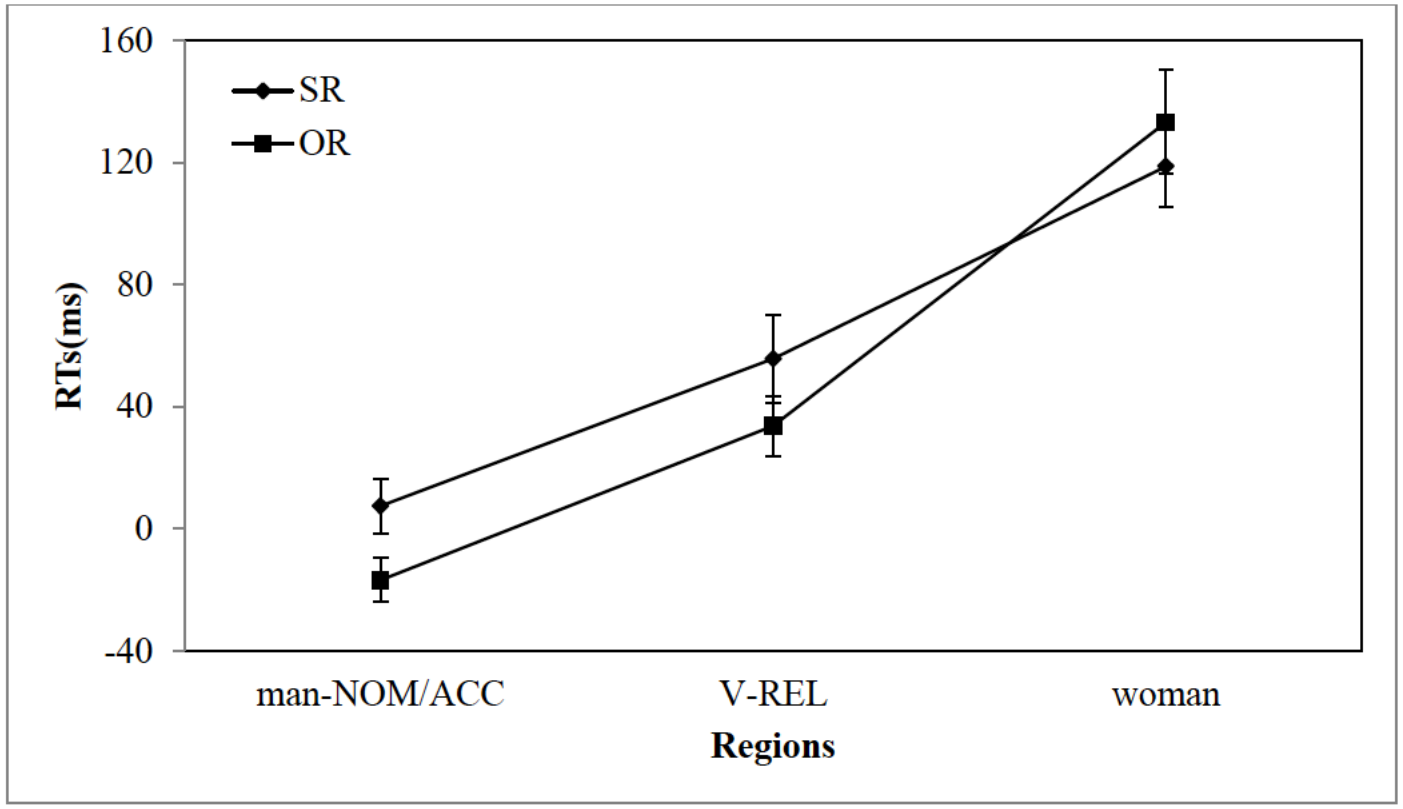



Figure1. Residual reading times per word
(Error bar indicates standard error)

# 2.4. Universal Processing Strategy?

- Syntactic factors play an important role in the process of English RCs, in addition to or more than pragmatic factors

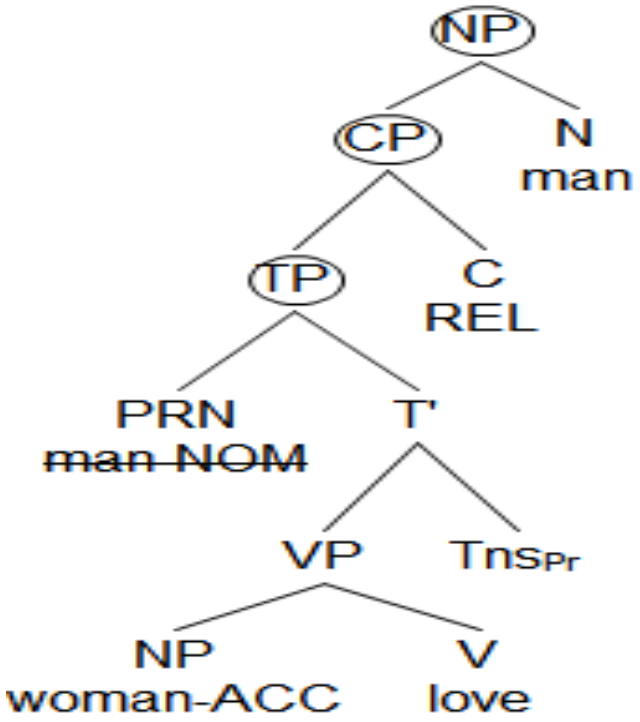
- However, in languages like Korean, Chinese, and Japanese, pragmatic factors may play a more prominent role.

# 2.4.1. Structural Distance Hypothesis

a. Subject relative clause

여자를 사랑하는 남자

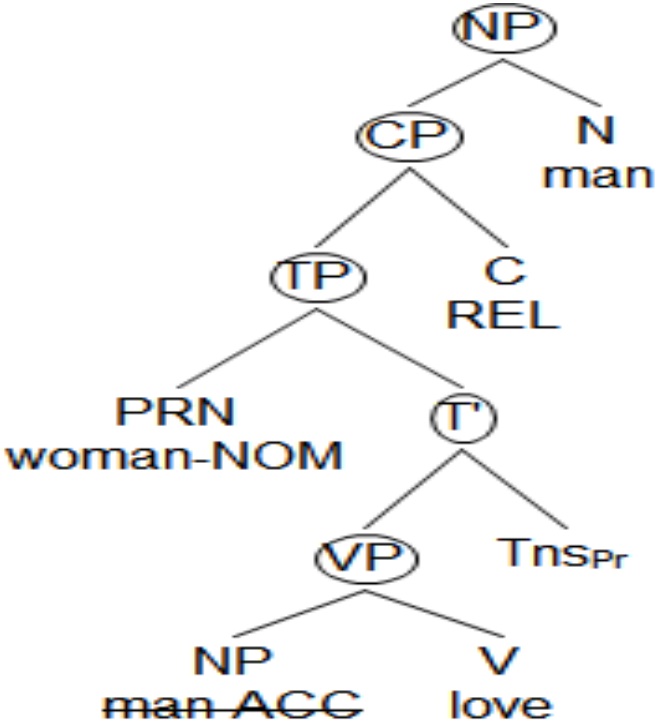
[_ yeca-lul salangha-nu-n] namca




b. Object relative clause

여자가 사랑하는 남자

[yeca-ka _ salangha-nu-n] namca



# Q1. Syntactic Distance in Korean RCs?

1) Null Anaphor Language (Chae, 2012)

- 철수는 ___ 믿는다.
- 그는 ___ 논리적이라고 생각한다.

# Q1. Syntactic Distance in Korean RCs?

2) Not Salient Island Effects Constraints? (Chae, 2012; Shin, 1996)

- Long-distance dependency such as RCs shows island effects.

  *He is someone [who nobody knows [what the FBA did to ~~who~~]].

  [[ ___ 사랑하던] 애인이 죽어버린] 철수

  [김교수가 [ ___ 수강하는] 학생들에게 모두 F를 준] 수업

# Q1. Syntactic Distance in Korean RCs?

3) Gapless RCs (Chae, 2012; Lee, 2012; Yeom, 2015)

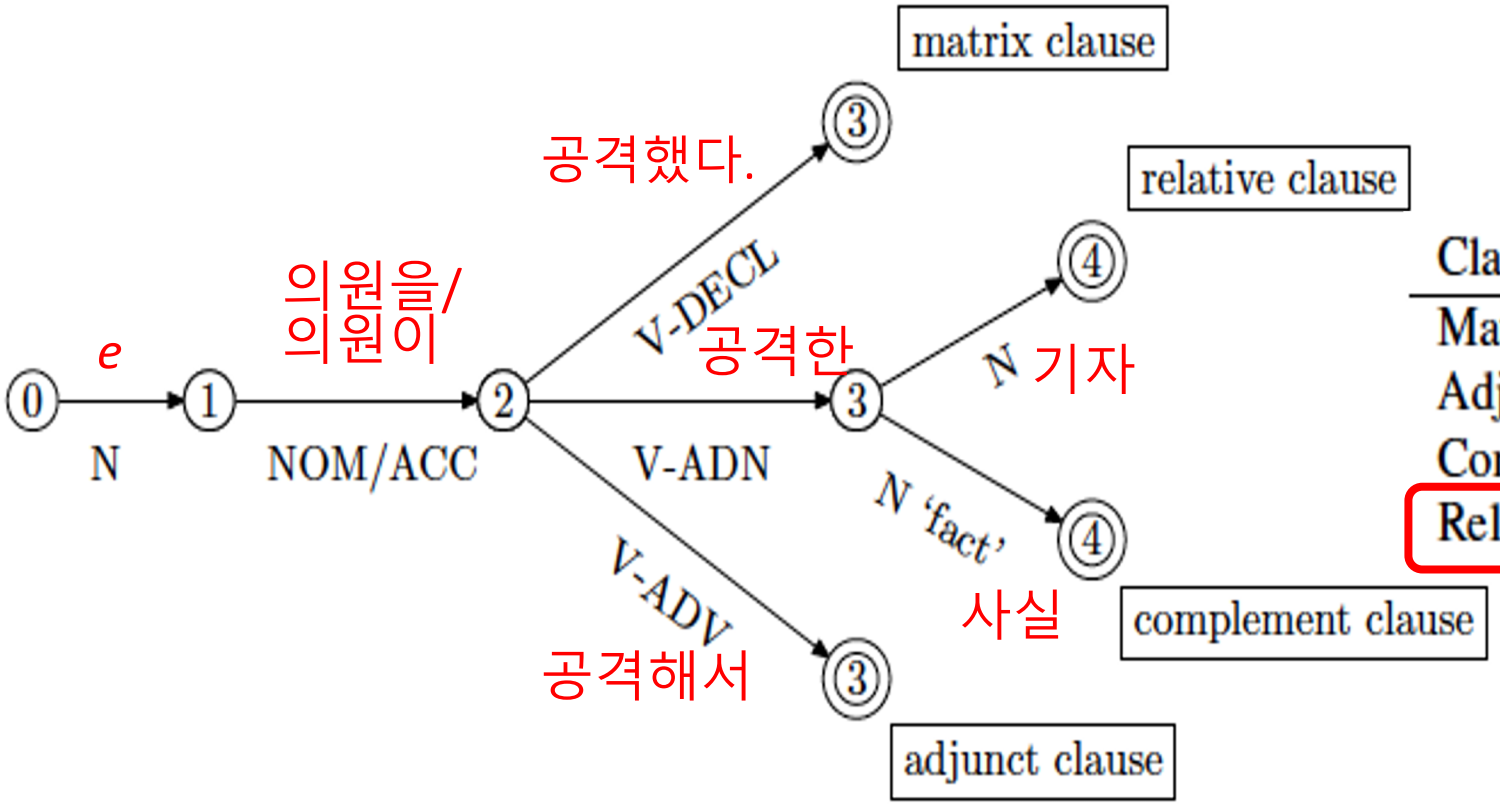
[생선이 타는] 냄새

[바람이 부는] 소리

→ RC as a noun-modifying construction (Comrie, 2003; Matsumoto, 1997)

# 2.4.2. Syntactic Probability Hypothesis

(3) Entropy in Korean




| Clause type | SBJ Extraction | OBJ Extraction |
|---|---|---|
| Matrix Clause | 19.6 | 19.6 |
| Adjunct Clause | 34.66 | 34.66 |
| Complement Clause | 32.1 | 42.98 |
| Relative Clause | 27.13 | 35.65 |

(Yun, Whitman, and Hale, 2010)

# Q2. Pragmatic Probability?

"[신문사의 사장이/을 비밀리에 정치적으로 이용한] 의원" (Kwon, 2008)

animate DP within RC — animate Head DP

Corpus study: Sejong Korean Parsed Written Corpus (28,434-*ejel*)

| | Animate DP within RC | | Inanimate DP within RC | |
|---|---|---|---|---|
| | Animate Head DP | Inanimate Head DP | Animate Head DP | Inanimate Head DP |
| SR | 11 | 8 | 119 | 99 |
| OR | 2 | 45 | 3 | 46 |

(KANG, 2014)

# Q2. Pragmatic Probability?

SR: [신문사 사장을 정치적으로 이용한] 의원/공금 (Kwon, 2008)

OR: [신문사 사장이 정치적으로 이용한] 의원/공금

animate DP within RC　　animate Head DP

| RCs | Animate DP within RCs | Head DP | |
|---|---|---|---|
| | | **Animate** (의원) | **Inanimate** (공금) |
| **SR** | 신문사 사장을 정치적으로 이용한… | **11** | **8** |
| **OR** | 신문사 사장이 정치적으로 이용한… | **2** | **45** |

(KANG, 2014)

# 2.5. Processing Implication

- Syntax factors may play a less important role in the processing of Korean RCs (than English RCs).

- Pragmatic interpretation (Choe, 2012; Yeon, 2012)
  [머리가 좋아지는] 책
  [연선생이 산] 백화점이 어디에요?

# 2.5. Processing Implication

Characterization Constraints (Kim, 2013)

- Korean relative clauses are subject to semantic, pragmatic constraints.
- Relative clauses must characterize, describe, or identify the head DP.
- In Korean, relative clauses are acceptable to the degree they satisfy the characterization constraint.

# 3. New Research Direction

- Can we provide more controlled, quantitative evidence that pragmatic factors play a more central role than syntactic ones in the processing of Korean RCs?
- What types of pragmatic factors influence the processing of Korean RCs? How can we model them?

# 3.1. Rational Speech Acts Model (RSA)

- Bayesian Inference : P $(H|D) = \frac{P(D|H) * P(H)}{P(D)}$

  Posterior probability: $P(H|D)$

  Prior probability: $P(H)$

  Likelihood: $P(D|H)$

- Posterior probability $\propto$ likelihood x prior probability

See e.g. Bergen et al. (2016)

# Strategic listener: $\boldsymbol{L_n}(w|u) \propto$ P($w$) $\boldsymbol{S_n}(u|w)$

- The listener interprets an utterance by integrating their **prior knowledge** with the likelihood that the speaker **Sn** would choose the utterance given different states of the world.

*u* = utterance
*w* = possible world
P(*w*) = prior probability of *w*
L(*w*|*u*) = probability that the listener assigns to *w* after hearing *u*.
S(*u*|*w*) = probability that the speaker will choose *u*
given the goal of communicating *w*.

# Strategic listener: $\boldsymbol{L_n}(w|u) \propto$ P($w$) $\boldsymbol{S_n}(u|w)$

- Ambiguous utterance: “연선생이 파는 백화점”

- Possible worlds

  1: 연선생이 백화점에서 (물건을) 파는 상황;

  2: 연선생이 백화점을 파는 상황

- Assumptions

  $S_n$(“연선생이 파는 백화점”|연선생이 백화점에서 파는 상황) = 1

  $S_n$(“연선생이 파는 백화점”|연선생이 백화점을 파는 상황) = 1

# Strategic listener: $L_n(w|u) \propto$ P$(w)$ $S_n(u|w)$

$L_n$ (연선생이 백화점에서 파는 상황|“연선생이 파는 백화점”)
$\propto$ P(연선생이 백화점에서 물건을 팔 확률)

$L_n$ (연선생이 백화점을 파는 상황|“연선생이 파는 백화점”)
$\propto$ P(연선생이 백화점을 팔 확률)

# 3.2. Experiment 1

- RQ1: Does a pragmatic factor affects the processing efficiency more than a syntactic factor in Korean RC?

- Probability of events — affects P(w)

  1) High probability

  경찰이 [범인이 바지 안에 숨기고 있다는] 사실을 알아채지 못한] **초소형** 권총

  2) Low probability

  경찰이 [범인이 바지 안에 숨기고 있다는] 사실을 알아채지 못한] **초대형** 권총

- Hypothesis: RCs characterizing an event that have high probability >

  RCs characterizing an event that have low probability

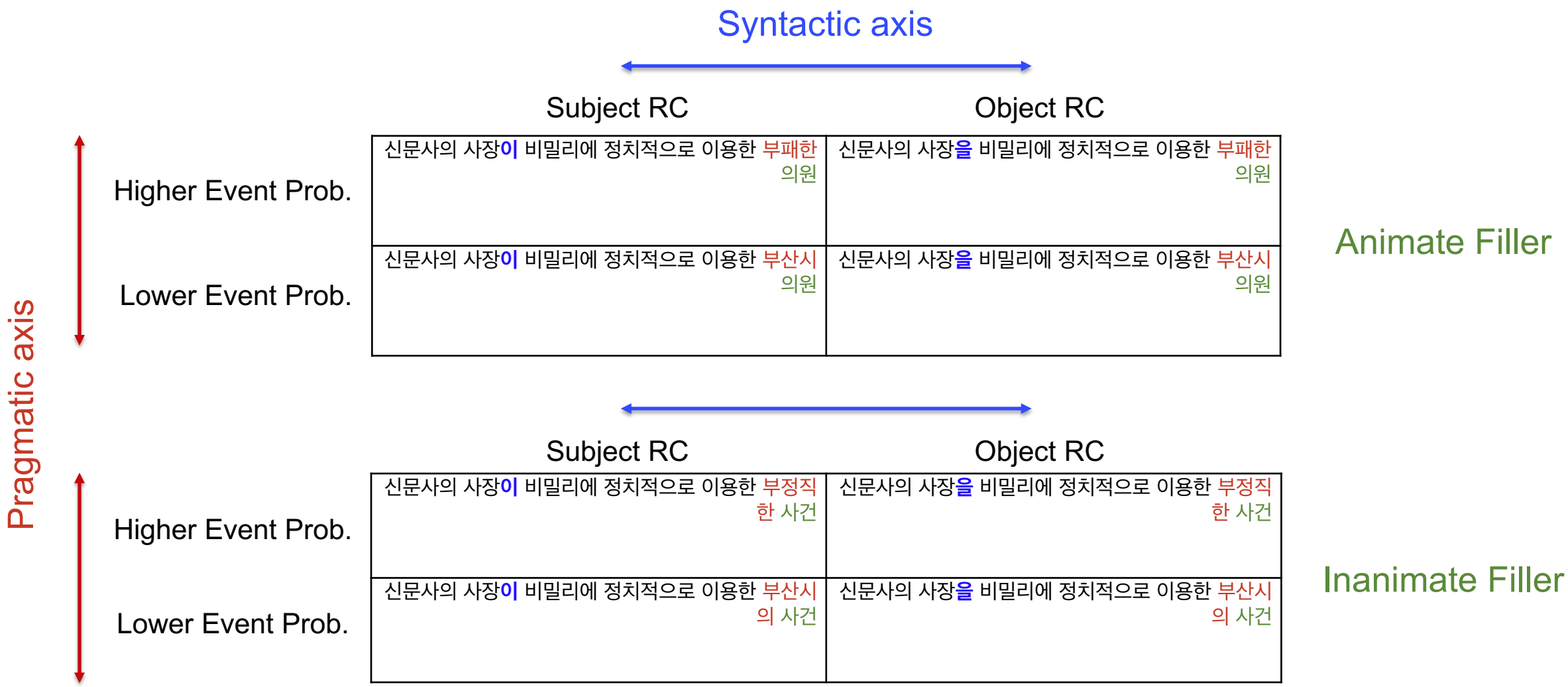
# 3.2. Experiment 1

- Method: Self-paced reading task while eye-tracking
- 2 × 2 × 2 factorial design

| Factor Name | RC type | Event probability — P(w) | Filler animacy |
|---|---|---|---|
| Level 1 | Subject RC | Higher | Animate |
| Level 2 | Object RC | Lower | Inanimate |

- Item * condition pairings allocation -> between-subjects design
- Independent measures: three factors
- Dependent measures: **RT difference** (over pragmatic/syntactic axes), comprehension accuracy, eye-path
- Predictions: RT difference bigger over pragmatic axis, different patterns depending on filler animacy
- Norming study to ensure the difference in the probability of events

# Sample Stimuli in Exp.1



# 3.3. Experiment 2

- RQ2: Does Korean speakers utilize different kinds of pragmatic knowledge in a systematic manner while processing ambiguous Korean RCs?

- Ambiguous Korean RC: “연선생이 파는 백화점”
- Prior knowledge — affects S(u|w)

  1) Biased Context:

  “연선생은 땅부자이다.” -> “연선생이 파는 백화점”

  2) Neutral Context:

  “연선생은 내 친구이다.” -> “연선생이 파는 백화점”

- Hypothesis: RCs characterizing a biased event >

  RCs characterizing an unbiased event

# 3.3. Experiment 2

- Method: Self-paced reading task while eye-tracking
- 2 × 2 × 2 factorial design

| Factor Name | Event probability — P(w) | Prior knowledge — S(u\|w) | Target sentence ambiguity |
|---|---|---|---|
| Level 1 | Higher | Biased | Ambiguous |
| Level 2 | Lower | Neutral | Unambiguous |

- Item * condition pairings allocation -> between-subjects design
- Independent measures: three factors
- Dependent measures: **query answer proportion**, RTs while reading, eye-path
- Predictions: effects of two pragmatic factors add up in a combinatorial manner, different patterns depending on target sentence ambiguity
- Norming study to ensure the difference in the probability of events

# Sample Stimuli in Exp.2

[Prior Knowledge]
① 예진이 / 말했다: / 연선생은 / 뭐를 / 해요?
② 수호가 / 말했다: / [내친구|땅부자] / 연선생은 / 이것저것 / 팔아요.

[Event Probability + Target Sentence Ambiguity]
③ 예진이 / 말했다: / 연선생이 / 파는 / [백화점|시장]이 / [**어디인지**|**얼마인지**]/ 알아요?
**어디인지** — [(백화점을) object, (백화점에서 (물건을)) oblique both possible]
**얼마인지** — [only (백화점을) object possible]
④ 수호가 / 말했다: / 잘 / 모르겠네요.

[Comprehension Query]
⑤ 연선생이 판 것은 [백화점|시장]이다. (Object reading) **vs**
연선생이 판 것은 다른 물건이다. (Oblique reading)

# Sample Stimuli in Exp.2

[Neutral Context]
① 예진이 / 말했다: / 연선생은 / 뭐를 / 해요?
② 수호가 / 말했다: / [내친구] / 연선생은 / 이것저것 / 팔아요.

[Lower Probability + Ambiguous]
③ 예진이 / 말했다: / 연선생이 / 파는 / [시장]이 / [**어디인지**]/ 알아요?

**>**

[Lower Probability + Unambiguous]
③ 예진이 / 말했다: / 연선생이 / 파는 / [시장]이 / [**얼마인지**]/ 알아요?

④ 수호가 / 말했다: / 잘 / 모르겠네요.

# Sample Stimuli in Exp.2

[Biased Context]
① 예진이 / 말했다: / 연선생은 / 뭐를 / 해요?
② 수호가 / 말했다: / [땅부자] / 연선생은 / 이것저것 / 팔아요.

[Higher Probability + Ambiguous]
③ 예진이 / 말했다: / 연선생이 / 파는 / [백화점]이 / [**어디인지**]/ 알아요?

**<**

[Higher Probability + Unambiguous]
③ 예진이 / 말했다: / 연선생이 / 파는 / [백화점]이 / [**얼마인지**]/ 알아요?

④ 수호가 / 말했다: / 잘 / 모르겠네요.

# Predictions in Exp.2

| Measure: Obj. > Obl. reading | Ambiguous target (어디인지) | | Unambiguous target (얼마인지) | |
|---|---|---|---|---|
| | Biased context (땅부자) | Neutral context (내친구) | Biased context (땅부자) | Neutral context (내친구) |
| Higher event probability (백화점) | 5 | 6 | 1 | 2 |
| Lower event probability (시장) | 7 | 8 | 3 | 4 |

| Measure: Processing difficulty | Ambiguous target (어디인지) | | Unambiguous target (얼마인지) | |
|---|---|---|---|---|
| | Biased context (땅부자) | Neutral context (내친구) | Biased context (땅부자) | Neutral context (내친구) |
| Higher event probability (백화점) | 1’ | 2’ | 1 | 2 |
| Lower event probability (시장) | 3’ | 4’ | 3 | 4 |

1’ < 1

4’ > 4

# 3.4. Further Directions

- Exp 2 in ERP: no semantic component (N400)

  [Biased Context]
  예진이 / 말했다: / 누가 / 팔아요?
  수호가 / 말했다: / 땅부자 / 연선생이 / 팔아요.

  [Target Sentences]
  예진이 / 말했다: / 연선생이 / 파는 / 백화점이 200억이에요.

# 3.4. Further Directions

- Exp 2 in ERP: Semantic component (N400)

  [Neutral Context]
  예진이 / 말했다: / 누가 / 팔아요?
  수호가 / 말했다: / 내친구 / 연선생이 / 팔아요.

  [Target Sentences]
  예진이 / 말했다: / 연선생이 / 파는 / 백화점이 200억이에요.

# 3.4. Further Directions

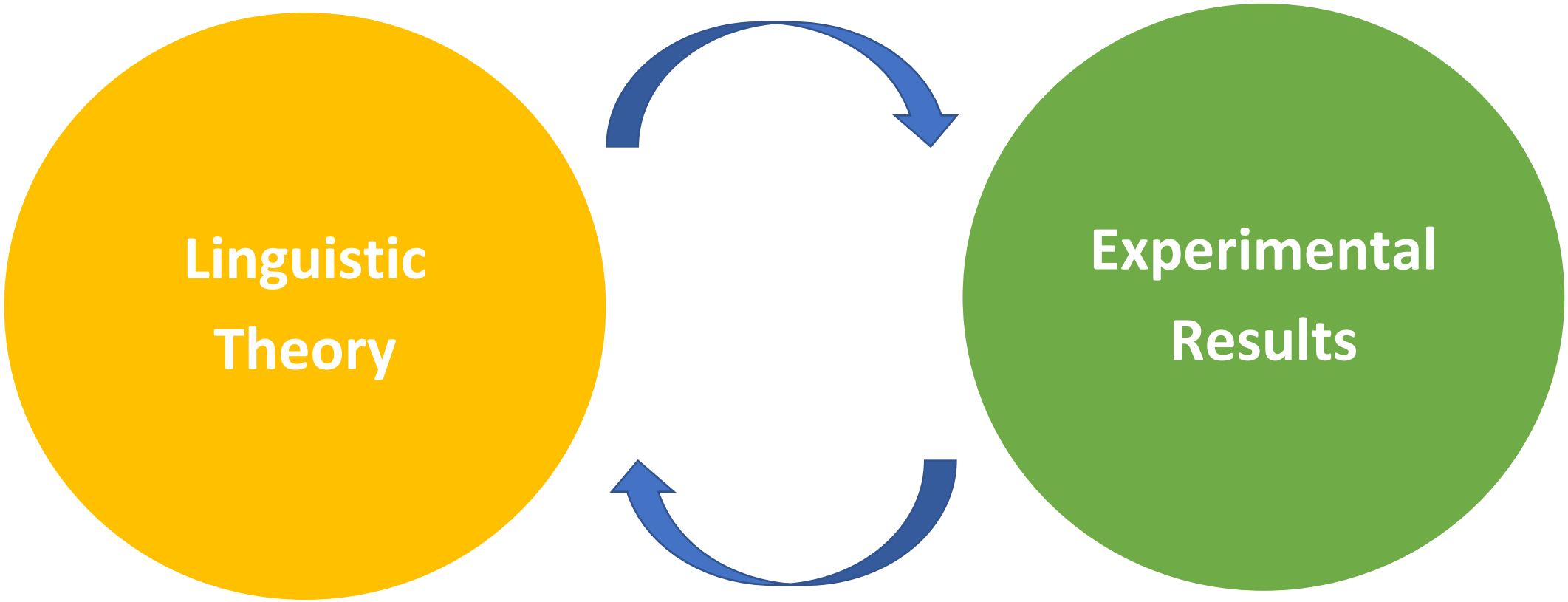
- Gapless Relative Clauses

  Gapless S:[고기를 맛있게 익힌] **냄새**가 요리사를 흥분시켰다.

  Gapless O: [고기가 맛있게 익은] **냄새**가 요리사를 흥분시켰다.

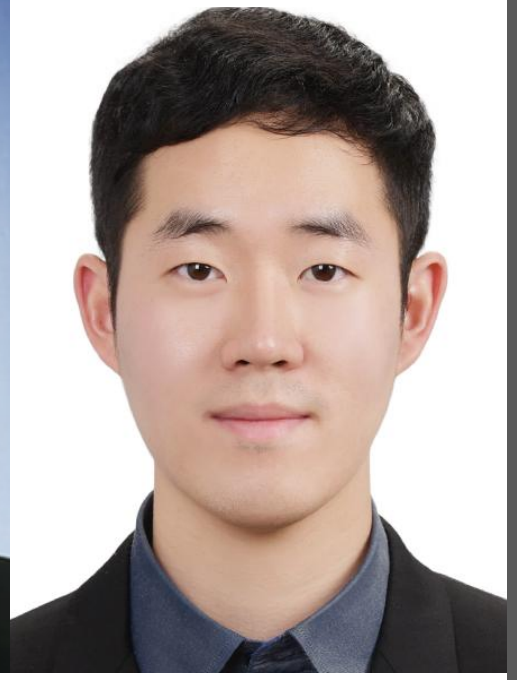
- Syntactic distance hypothesis cannot explain longer RTs at Head DP whereas pragmatic hypothesis may predict assymetry in RTs at Head DP.

Guide for language processing




Feedback into theoretical linguistics

# Project Members



## Meaning & Melody Lab

- PI: Sunwoo Jeong
- Soyeon Kang
- Gyuhwan Lee

# References


Bergen, L., Levy, R., & Goodman, N. (2016). Pragmatic reasoning through semantic inference. Semantics & Pragmatics 9, 20.

Chae, H. (2012). Are there relative clauses in Korean? A participle clause analysis. *Korean Journal of Linguistics*, *37*(4), 1043–1065.

Gibson, E. (2000). The dependency locality theory: A distance-based theory of linguistic complexity. In A. Marantz, Y. Miyashita & W. O'Neil (Eds.), Image, language, brain : Papers from the first mind articulation project symposium. 95-126. Cambridge, MA: MIT Press.

Hale, J. (2006). Uncertainty about the rest of the sentence. *Cognitive Science*, *30*(1), 1-30.

Keenan, E., & Comrie, B. (1977). Noun phrase accessibility and universal grammar, *Linguistic Inquiry*, *8*(1), 63-99.

Kim, I. (2013). Rethinking “island effects” in Korean relativization. *Language Sciences*, *38*(1), 59-82.

Kwon, N. (2008). Processing of syntactic and anaphoric gap-filler dependencies in Korean: Evidence from self-paced reading time, ERP and eye-tracking experiments. Ph.D. dissertation. University of California, San Diego.

Lee, C., & Lee. J. (2012). Gap in “Gapless” relative clauses in Korean and other Asian languages. UCLA Working Papers in Linguistics, *17*(25), 204-214.

Levy, R. (2008). Expectation-based syntactic comprehension. *Cognition*, *106*(3), 1126-1177.

# References


O'Grady, W. (1997). Syntactic development. Chicago, IL: University of Chicago Press.

Shin, H. (1996). The syntax and semantics of Korean relative constructions. PhD Dissertation, Seoul National University.

Sprouse, J., I. Caponigro, C. Greco, and C. Cecchetto (2016). Experimental syntax and the variation of island effects in English and Italian. Natural Language & Linguistic Theory 34(1), 307–344.

Yeom, J. (2015). Gapless adnominal clauses in Korean and their interpretations. *Language Research*, *51*(3), 597-627.

Yeon, J. (2012) 'A functional-typological study on Korean relative clauses in Korean.' *Journal of Korean Linguistics*, *63*(1), 413-457.

Yun, J., Whitman, J., & Hale, J. (2010). Subject-object asymmetries in Korean sentence comprehension. *Proceedings of the Annual Meeting of the Cognitive Science Society*, *32*(1).